



ARTICLE




<https://doi.org/10.1057/s41599-023-02096-w>

OPEN

# General theory of data, artificial intelligence and governance

Pablo de Pedraza <sup>1,2</sup>  & Ian Vollbracht<sup>1</sup>

Big Data (BD) and Artificial Intelligence (AI) play a fundamental role in today's economy that traditional economic models fail to capture. This paper presents a theoretical conceptualisation of the data economy and derives implications for digital governance and data policies. It defines a hypothetical data-intensive economy where data are the main input of AI and in which the amount of knowledge generated is below the socially desired amount. Intervention could consist of favouring the creation of additional knowledge via data sharing. We show that the framework suggested describes many features of today's data-intensive economy and provides a tool to assist academic, policy and governance discussions. Our conclusions support data sharing as a way of increasing knowledge production on societal challenges and dilemmas of data capitalism and transparency in AI.

<sup>1</sup>European Commission, JRC, Ispra, Italy. <sup>2</sup>University of Salamanca, Salamanca, Spain. email: [pablodepedraza@gmail.com](mailto:pablodepedraza@gmail.com)

## Introduction

The ability to transform Big Data (BD) into products and services is a game-changing factor for the whole economy that is setting challenges for society, such as concentration of power and lack of transparency. From an economic point of view, there is a need for new theoretical developments because the traditional price-quantity approach has limitations in capturing the workings of the data-intensive economy (Khan, 2017). In this paper we use a new theoretical framework to look at data capitalism through the lens of economic principles and derive implications for digital governance data policies.

In data capitalism, Big Data and Artificial Intelligence (AI) are used intensively to produce knowledge and services. Given the central role of data as the main AI input, our initial research question is whether economic theory supports data sharing as a fair, equitable, *inclusive* policy (European Commission, 2018a, European Commission, 2020, Acemoglu and Robinson, 2012, Galenson, 2017, Chhillar and Aguilera, 2022). We find theoretical grounds for data sharing, then explore how it can tackle existing societal challenges while taking into account governance dilemmas.

We begin by using the circular flow model and a country-level macroeconomic equilibrium equation (see Annex 1, Samuelson 1948, Samuelson and Nordhaus 2010) to point out two issues. On the one hand, the data-intensive economy generates disequilibria in the traditional equation through both leakages (taxes, saving and imports) and injections (government spending, investment and exports). Current governance based on fines levied by courts can help to solve conventional disequilibria (Singer and Isaac, 2020, Stempel, 2020, Chhillar and Aguilera, 2022). However, the traditional equation does not capture the societal challenges in the data and knowledge dimensions of the economy, which motivates further theoretical developments.

We build upon the circular flow model (Annex 1) to present a theoretical framework that augments the traditional view of the economy for goods and services by adding Big Data, AI and knowledge flows. Our theoretical conceptualisation builds upon seven assumptions. First, daily activity by households and firms generates BD. Second, data holders collect BD to produce knowledge using AI and utilise/monetise it in the form of services. Third, data are a means of payment which challenges traditional thinking on prices and quantities. Fourth, in the initial stages of digitalisation consumers maximise their utility, assuming that the monetary value of their individual data is close to 0. This assumption means that when individuals pay for a service with their personal data they consider the service to be 'free'. Fifth, data holders are profit maximisation agents who treat data as a valuable asset. Sixth, the value of data increases when they can be an input in the production of services that show network effects. Seventh, decreasing returns to scale in knowledge production from BD using AI only appear when  $N = \text{All}$  and  $X = \text{everything}$ . We represent this economy using the semi-circular flow of the economy diagram.

We continue the analogy by defining *data leakage*, *knowledge injection* and their corresponding inclusive policies. A *Data Sharing (DS) Policy* consists of removing barriers to data access (*leakage*) to generate additional knowledge (*injection*) about societal challenges. We take into account governance dilemmas (Chhillar and Aguilera, 2022) by assuming that *DS* operates in a similar way to monetary taxation and specify and draw a *Data Sharing Laffer Curve*: this shows the theoretical relationship between the data sharing rate and the amount of knowledge generated.

Next, we study the extent to which our theoretical model captures reality. We show that it is in line with existing evidence showing that the amount of knowledge disclosed in on-line

markets tends towards monopoly levels (Board and Lu, 2018), and with several streams of literature such as anti-trust (Khan, 2017, Cr mer et al., 2019), intangible assets (Govindarajan et al. 2018), Big Data Business Models (BDBM) (Wiener et al. 2020) and recent literature reviews on AI Governance and management (Chhillar and Aguilera, 2022). Our conclusions support several proposals already put forward: the World Economic Forum multiple stakeholders approach, the establishment of a data authority (Scott Morton et al. 2019, Martens, 2016) and further implementation and development of data portability rights (De Hert et al., 2018, European Union, 2016) by means of Personal Data Stores (PDSs) (Bolychevsky and Worthington, 2018). They also support the United Nations' call for a global partnership to improve the quality of statistics available to citizens and governments to reduce gaps between the private and public sectors (UN, 2013, 2014, OECD 2016, WEF 2019).

Finally, the Discussion section and our main takeaways build on the literature on Digital Governance by offering examples of data sharing that would generate knowledge about societal challenges identified by Chhillar and Aguilera (2022).

The semi-circular flow of the economy is a development of the circular model that underlines the usefulness of economic thinking in offering an overview of the data economy that links some of the more difficult and less tractable elements. It is a theoretical contribution that provides a framework for reducing the conceptual complexity of the digital economy. It is consistent with the existing literature in several research lines. The theory enhances value creation from BD and AI while identifying its social costs and Governance challenges. We conclude that fostering data sharing (data leakages) in the real world would incentivise research and generate useful knowledge about the societal challenges of data capitalism and its governance dilemmas, fostering innovation and technology and spreading the economic benefits of the digital revolution (knowledge injections).

We refer to the work presented here as a general theory because it captures the functioning of data capitalism in the same way that the traditional circular model captures the workings of the capitalist system prior to the advent of BD and AI. We do not claim to have developed a full system that covers every uncertainty on the topic but rather a framework able to facilitate multidisciplinary academic and public-policy discussions, their interactions in this area and further theoretical developments. Our paper gives strong arguments to policy-makers and researchers in claiming better data access. Although the traditional government intervention rationale applies to a data intensive economy, our framework also has some caveats, which we review in the Discussion section.

The rest of the paper is organised as follows. Section 'A theoretical framework for a data intensive economy governance' conceptualises a hypothetical data-intensive economy, its sources of disequilibrium, its market failures and its governance by means of data sharing to increase knowledge. Section 'Does the world fit the semi-circular model?' explores how the real world fits the semi-circular model conceptualisation. Section 'Discussion: How? What? Who?' discusses how to activate data leakages and what a Pareto-efficient data sharing policy would look like. Section 'Conclusions' concludes. Two annexes summarise the traditional circular flow model and the traditional Laffer curve.

## A theoretical framework for a data intensive economy governance

Throughout the paper, we refer to BD, knowledge, AI, production of services and data holders.

BD and knowledge are intangible assets that can be transformed into profitable products and services. Following the DIKW (Data, Information, Knowledge, Wisdom) hierarchy, data are elementary, recorded descriptions of things, events, activities or transactions which are unorganised and unprocessed. Knowledge is data organised and processed to convey understanding of a current problem, enabling action (Rowley, 2007) and production. The better the data represent the real-world construct to which they refer, the more activities and agents they capture, the better the quality of BD is. The quality and value of data also depend on how useful they are for a specific purpose (Redman, 2013), such as profit and utility maximisation. Data value is released when data are shaped and organised for the specific purpose of production of services (Redman, 2018, Varian, 2018).

The important role of knowledge as a particular form of information is not a new concept in economics. It has traditionally played a fundamental role in the market economy and in defining what role is appropriate for governments (Stiglitz, 2001). Knowledge is a key component in productivity and growth (Romer, 1986), the *one ring of globalisation* that rules trade, capital flows and immigration (Freeman, 2013). In a data-intensive economy, access to BD and AI determines knowledge creation, value capture (Chhillar and Aguilera, 2022), welfare (Duch-Brown, 2017a, 2017b, and 2017c), innovation, wealth and power distribution (OECD, 2019, ITU, 2018).

When we refer to AI, we mean a scaled-up automated application of existing statistical techniques that enables patterns, regularities and structures in data to be recognised without an a priori theoretical framework (Boisot and Canals, 2004, Duch-Brown et al. 2017, Vigo, 2013, Duch-Brown, 2017a, 2017b, 2017c). We therefore take a very broad definition of AI covering machine learning and related methods that can be used to analyse BD in order to generate knowledge that enables services to be produced and value to be captured. BD are data characterised by their volume, velocity and variety (Laney, 2001, 2012). Massive numbers of data points can be collected, organised, combined, searched and used for a wide variety of analysis purposes. AI models can be tested and continuously improved with new BD. Algorithms trained on one data set can be transposed to other complementary data sets and adjacent data (Duch-Brown et al., 2017, Duch-Brown, 2017a, 2017b, 2017c) to obtain more and better predictions. We consider that AI is data-driven and that data facilitate AI advancement.

'Data Holders' here means born-digital companies that operate globally. In terms of the literature on BDBM, data holders are vertically integrated. They create and capture value by internalising the whole BD life-cycle from data collection to analysis and use, including aggregation and analyses for strategic decisions and internal operations, plus product enrichment. Their vertical structure means that data holders have the necessary infrastructure to be users, suppliers and facilitators at the same time (Wiener et al. 2020). This also implies that they incur high fixed costs.

**Disequilibria in a data-intensive economy.** The circular flow of the economy represents the macro-level exchanges in pre-data capitalism where money is the unique means of payment. It represents the traditional disequilibria and the corresponding government interventions via monetary leakages and injections (see Annex 1). Based on the traditional circular flow of the economy, Eq. (1) captures the traditional country-level macro-economic equilibrium. The state of (macro) economic equilibrium occurs when total leakages (savings ( $S$ ) + taxes ( $T$ ) + imports ( $M$ )) are equal to total injections (investment

( $I$ ) + government spending ( $G$ ) + exports ( $X$ )) in the economy. This can be represented by:

$$S + T + M = I + G + X \quad (1)$$

Disequilibrium occurs when leakages are not equal to total injections. In such a situation, changes in expenditure and output will lead the economy back to equilibrium. Such changes will depend on the type of inequality ( $S + T + M > I + G + X$  or  $S + T + M < I + G + X$ ).

Equation 1 describes the macro-economic flows in a pre-data open economy. It captures certain aspects of the data-intensive economy. For example, data holders are typically supra-state agents that operate globally and frequently concentrate in low-tax jurisdictions, which decreases the ability of governments to collect taxes ( $T \downarrow$ ), which in turn reduces governments' financial capacity, spending ( $G \downarrow$ ) and ability to respond to market failures, promote efficiency, equity and stability. At the same time, data holders' ability to collect valuable data increases their financial power and ability to attract investment ( $I \uparrow$ ). Data and the ability to process them are a critical ingredient of innovation, knowledge and value creation, which makes data holders an attractive store of value for investors. The resulting financial strength often allows data holders to predate markets (Khan, 2017). Creative destruction applies also to sectors traditionally provided by the State such as health, education, public transport, currencies and national defence<sup>1</sup>.

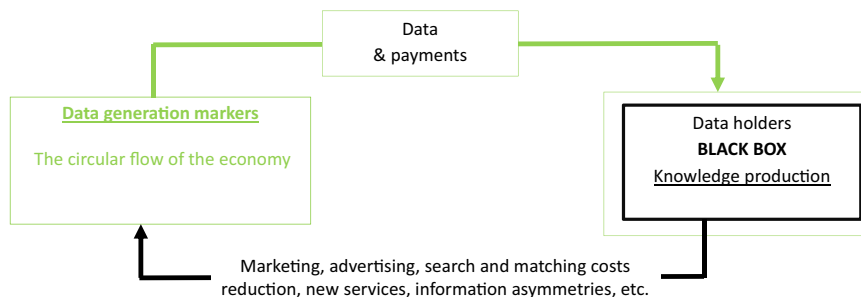
Current governance tackles these disequilibria via fines levied by courts. Monetary taxation of digital activities and fines may help to balance Eq. (1) without the need to reduce government expenditure ( $G \downarrow$ ), but it does not tackle the data and knowledge aspects of a data-intensive economy. As reported in the literature, fines levied by traditional governance does not help to provide an understanding of data capitalism paradoxes, or to correct societal challenges such as power imbalances, opacity and unfair distributions in value capture (Singer and Isaac, 2020, Stempel, 2020, Chhillar and Aguilera, 2022, Khan, 2017 and Lehdonvirta, 2022). In the next section we draw up a new theoretical framework that incorporates the data and knowledge dimensions of data capitalism<sup>2</sup>.

**Seven assumptions for a data-intensive economy.** Figure 1 represents a data-intensive economy characterised by the following seven assumptions, which are a theoretical representation of data capitalism.

First, on the left-hand side, households and firms operate according to the 'circular flow of the economy' model (see Annex 1), exchanging goods and services for money and labour for wages, generating a circular flow of money (Samuelson, 1948, Samuelson and Nordhaus, 2010). Their activity—in parallel—generates a flow of data towards data holders, who are represented on the right hand side. This is not captured by the traditional model.

Second, on the right-hand side, data holders use BD and AI techniques to extract knowledge from data to produce digital services. Knowledge production generates new, innovative services that influence the left-hand-side markets by matching efficiency, marketing, advertising and reducing search and transaction costs. Data flows are semi-circular: from households and firms<sup>3</sup> to data holders but not in the other direction. Households and firms receive data-driven services created by data holders based in part on their own data, but do not receive unprocessed data. This a fundamentally semi-circular rather than circular relationship.

The additions to the traditional circular flow of the economy (see Annex 1) so far are the prominent role of data flows, data



**Fig. 1 The semicircular flow of the data economy.** Provision of digital services (black arrow) in return for data (green arrow).

holders and knowledge and service production from data on consumers. These first two assumptions, which together characterise the data economy as semi-circular, stress that data flows only in one direction: from households and firms to data holders. The semi-circular assumption does not neglect the fact that there is an exchange between households and firms and data holders. The fundamental distinction from the standard circular flow model is the following: in the circular model, both the prices that consumers/workers pay for goods and services and the wages that consumers/workers receive for the labour that they supply have explicit monetary prices. However, in the semi-circular economy the latter are characterised by a barter exchange: namely, provision of (free) digital services in return for data, but at an unknown and not explicitly defined rate of exchange. This is the fundamental difference from the traditional model where both sides of the market operate with explicit monetary prices. This leads to the third assumption.

Third, data are a means of payment by firms and families to data holders. In the circular flow of the economy explicit prices are a fundamental variable, but data are ambiguous as a means of payment. Money is easy to use and understand, but data are not. Data are not easily priced and their value is not clear, especially at the individual level. Data flows do not generate clearly comparable market signals in the way that prices do. There is also no authority in charge of setting aggregate data value comparable to the role of central banks that set interest rates to regulate the money supply to the wider economy. This challenges the traditional 'prices and quantities' thinking of standard economics. A space where prices are paid in data and quantities refer to digital services cannot be drawn in such a simple manner, but economic principles such as utility and profit maximisation, market failures and government intervention still apply to it.

Fourth, from a micro-economic point of view, consumers' utility maximisation is a function of monetary prices (MP), individual data value (IDP) and quantity (Q). In the early stages of digitalisation most consumers very often fail to realise that they are generating data, ignore terms and conditions of data transfer when using on-line services, and behave as if the monetary value of their personal data were effectively zero ( $IDP = 0$ ). Thus, if the price of a digital service is paid only in personal data, consumers consider it a 'free' service and give away their data 'for free' in monetary terms, in a barter exchange for a specific functionality. As a result, consumers and many companies maximise their utility considering only the explicit monetary part of the prices applied.

This assumption does not necessarily mean that consumers prefer to share their data freely and end up losing monetary value. Consumers and many companies are unable to capture any value from their data other than that which they obtain when using the digital functionality in the barter exchange. In fact, even if the economic value of individual data is positive before it is merged with other data from other individuals, its value is indeed

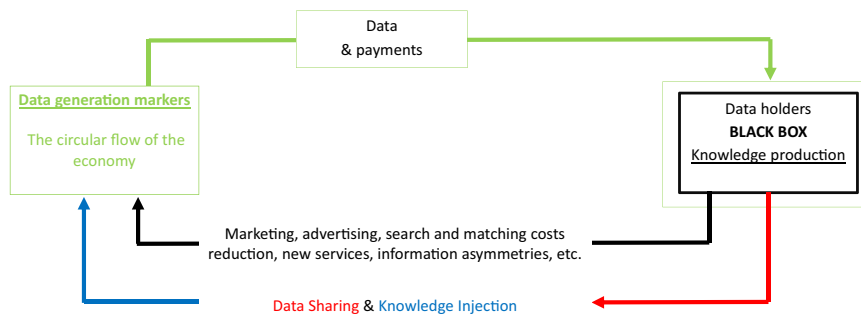
typically close to zero in isolation. The ability to create innovative services comes from the integration of data from many individuals. Individuals own their personal data, but their value can only be captured after data holders have merged and processed them and produced services with them. Even if there is a clear legal corpus assigning individuals ownership of their personal data, and a data flow back to them, the average individual has no ability to process them and no knowledge of possible alternative uses of data, so individuals act as if there is no opportunity cost of giving away their data for 'free'.

However, digitalisation is a dynamic process and the  $IDP = 0$  assumption holds in the specific circumstances of the initial stages of data capitalism if data holders' activities within a black box do not undermine rights, such as privacy, or have any negative impact on the functioning of competitive markets and the rule of law. Under these circumstances, individuals do not, in general, consider their personal data to constitute a valuable asset. The zero data value assumption fails however, hence  $IDP > 0$ , if individuals are aware that payment using data entails a cost in terms of individual rights such as privacy, an opportunity cost or the foregoing of income as a result of providing for free data which in fact have a close-to-zero but nonetheless positive market value.

Fifth, data holders are profit maximisation agents. They obtain 'de facto' ownership of data, build hugely valuable BD sets, draw and capture value from them by extracting knowledge to produce digital services. They consider data as a valuable asset: they store them, put them to work and manage them appropriately to obtain value. As profit maximisation agents, they set barriers to data access and generate the amount of knowledge and services that maximises their profit and fosters their market power.

Sixth, data generation markets show both direct and indirect network effects: the numbers of users on both sides of the market increase the value of the service. Data holders seek to increase their market power by expanding network effects to as many activities as possible. Data quality and value increase with their ability to foster network effects: the more activities and agents are covered by data, the better those data represent the real-world construct and the more valuable they are for producing services that expand network effects across human activities.

Seventh, in knowledge production that seeks to reach as many aspects of life as possible, efficiencies arise from volume (scale) and variety (scope) and average costs become lower as the data set grows bigger. Data holders and investors are in a race towards bigger, more detailed data sets, towards  $N = \text{all}$  and  $X = \text{everything}$ . In statistical terms, scale refers to the number of observations ( $N$ ) and scope to the number of explanatory variables ( $X$ ). Volume helps specify models because the larger the number of individuals observed ( $N$ ), the greater the degrees of freedom to include more variables ( $X$ ). The opposite also applies: the higher the number of variables, the bigger the sample that is needed. Scale and scope reinforce each other and are a direct consequence of the two Vs in



**Fig. 2 The semicircular flow of the data economy.** A data sharing policy (leakage: red arrow) enables additional knowledge to be produced (injection: blue arrow).

the definition of BD (Laney, 2001, 2012): volume and variety. Knowledge extraction from BD using AI has high fixed costs and almost negligible variable costs. It is unknown when diminishing returns to scale in knowledge extraction from BD appear. In a simple ordinary least squares estimation of one dependent variable as a function of several explanatory ones, diminishing returns may appear after a few thousand observations. After a certain  $N$  randomly extracted from the same population, estimated elasticities change very little. That is not necessarily the case if knowledge seeks to expand to as many aspects (Xs) as possible of individuals' lives. In that case, economies of scale and scope operate together, reinforce each other and operate with network effects. This makes massive, detailed data sets very valuable even if  $IDP=0$  or is very close to zero for each individual data point.

Concentration arises from the interaction between network effects and the efficiencies that derive from lowering the average cost of collecting data and producing knowledge to deliver services that display and foster network effects. According to natural monopoly logic, a small number of operators have incentives to collude<sup>4</sup> and the relevant market structure may tend towards a single operator over time. To take an extreme hypothetical case, where AI technology is common across all firms, access to the largest BD 'lake' enhances efficiency. This creates a 'winner takes all' dynamic in which the holder of the largest BD lake generates the largest knowledge rents and the most innovation, which can then be used in part to create new services and network effects, to further augment the size and detail of the BD 'lake', thus creating a self-reinforcing loop.

The fifth, sixth and seventh axioms imply that in a data-intensive economy knowledge production using AI and BD tends towards a natural monopoly where network effects, economies of scale and scope, high fixed costs and other barriers to entry operate together. Oligopolistic data holders compete but have incentives to collude and end up as a monopoly. The amount of knowledge produced, its disclosure, its prices and its quantities then tends towards the implications of monopoly theory (Schumpeter, 1942):

- The expansion of data holders across sectors and activities generates a process of *creative destruction* that replaces less efficient, less effective traditional operators that lag behind in their ability to collect BD and generate knowledge.
- Market structure and lack of competition *attract investment* for R&D and innovation.
- *Data holders are able to set prices and quantities for the services that they provide.* Regarding prices paid in data, the market works on a 'take it or leave it' basis: digital services are often only available to consumers who are willing to provide data as an (implicit) part of the bargain. Regarding quantities, data holders set the amount of knowledge

production at the amount that maximises their profit. That quantity is below the socially desirable amount. In practice, controlling knowledge production and its disclosure means that there are information asymmetries between data holders and the rest of the agents in the economy such as consumers, relevant government agencies (such as central banks and antitrust authorities) and the scientific research community. Barriers to entry mean that 'de facto ownership' by data holders keeps citizens and the public sector outside the black box and denies them access to the data lake.

As in a monopoly, the amount of knowledge produced is below perfect competition levels. Under-production of knowledge entails an opportunity cost for society as a whole. Some knowledge is not produced, so society loses it. Part is produced but not disclosed, and is therefore subtracted from the consumer surplus in a hypothetical BD-knowledge space.

**A data policy.** In this quasi-monopolistic framework, how can governance foster an increase in knowledge production and its disclosure to the public? If the seven assumptions above hold, in a data intensive economy knowledge production using BD and AI tends towards a natural monopoly, which leads to *under-production of knowledge*. Following the 'semi-circular flow of the economy' analogy, Fig. 2 represents a data policy in which new data flows are generated (leakage<sup>5</sup>) to enable additional knowledge to be produced (injection). We refer to this type of data policy as *Data Sharing Policy*. It consists of removing barriers to data access to generate additional knowledge on, for example, how to meet traditional government goals of promoting efficiency, equity and stability in the context of societal challenges and market failures involving data capitalism.

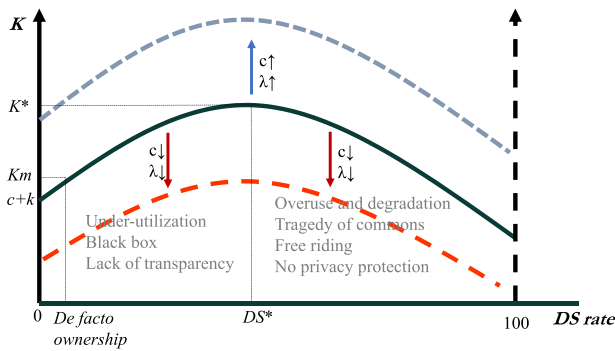
Assuming that data leakages (red arrow) operate as monetary taxation, Eq. (2)<sup>6</sup> identifies the determinants of the amount of data shared. We refer to this as *Data Revenue (DRev)*:

$$DRev = DS * DB \tag{2}$$

where  $DB$  is the *Data-Sharing Base* (the total amount of data produced by data holders) and the  $DS$  is the *Data-Sharing Rate* (the percentage of data produced by data holders that is shared). As  $DS$  changes,  $DRev$  changes as follows:

$$\frac{\partial DRev}{\partial DS} = DS * DB + DS * \frac{\partial DB}{\partial DS} \tag{3}$$

The two terms on the right-hand side describe two different effects. The first term is the direct effect of a  $DS$  rate increase on  $DRev$ . If the total  $DB$  were perfectly inelastic with respect to  $DS$ , the first term would be the only effect. With no behavioural response from households, firms and data holders,  $DRev$  would simply increase proportionately to changes in the  $DS$ . However,



**Fig. 3** The theoretical representation of the data sharing Laffer curve.  $DS^*$  is the  $DS$  rate that maximises  $K$ . Changes in  $\lambda$  (technology) and/or  $c$  (trust, stability and corruption) move the curve upwards or downwards.

economic agents may respond in several ways to changes in  $DS$ . As happens with taxes, as the  $DS$  increases data holders may undertake fewer activities that generate data, may increase data sharing avoidance activities and may shift their activities to countries where the  $DS$  is lower. Households and firms may also change their data generation behaviour. Thus, the overall effect of  $DS$  changes on  $DR$  is ambiguous. A higher  $DS$  increases  $DR_{rev}$  by the first term but may increase or decrease by the second.

Regarding knowledge injection (blue arrow), Eq. (3) separates the determinants of Knowledge Injection ( $K$ ):

$$K = (\lambda * DS * DB) + k + c \tag{4}$$

where  $\lambda$  refers to technology and the ability to process BD and obtain meaningful conclusions;  $k$  to knowledge generation that is independent of BD and AI; and  $c$  to idiosyncratic country-specific characteristics and circumstances such as trust, stability and corruption. As  $DS$  changes Knowledge Injection ( $K$ ) changes as follows:

$$\frac{\partial K}{\partial DS} = \lambda * DS * DB + DS * \frac{\partial DB}{\partial DS} * \lambda + DS * \frac{\partial \lambda}{\partial DS} * DB \tag{5}$$

where the first two terms on the right-hand side refer to the direct, and behavioural, effects described above and the third refers to how technology and knowledge generation ability change with changes in the  $DS$ .

If behavioural reactions to increases in  $DS$  reproduce those of traditional taxation,  $DS$  is unlikely to affect  $DR_{rev}$  and  $K$  linearly. We postulate Eq. 6 as a plausible functional form for describing the relationship between  $K$  and  $DS$ . It assumes that the relationship is positive at low  $DS$  rates, and that increases in the  $DS$  increase  $K$ . At higher tax rates the relationship is negative and the elasticity of  $DB$  with respect to  $DS$  rate ( $\beta_3$ ) may exceed one, causing knowledge generation to fall as the  $DS$  increases.

$$K = c + k + \beta_1 \lambda + \beta_2 DS - \beta_3 DS^2 \tag{6}$$

The graph, where the horizontal axis represents the  $DS$  and the vertical axis represents  $K$ , resembles the traditional Laffer curve (see Annex 2). We refer to it as the Data Sharing Laffer Curve (Fig. 3)<sup>7</sup>.

The  $DS$  rate is a number between 0 and 100. ‘ $DS = 0$ ’ represents an economy with no data-sharing responsibilities. The only knowledge generated is independent from AI and BD. From  $DS = 0$ , data holders start drawing value from BD to maximise their profit and their market power. Data holders can even voluntarily start opening up BD, making them available to other agents by means of APIs or ad hoc non-disclosure agreements. Data holders’ ‘data philanthropy’, marketing and willingness to activate a research community around their interests are represented as  $DS$  rate = ‘de facto ownership’, which generates

more knowledge than at  $DS = 0$ . From  $DS =$  ‘de facto ownership’ governments, in the exercise of their monopoly on power, can increase data-sharing pressure depending on different political views of efficiency, equity, stability, privacy, surveillance, market failures and Data Capitalism Challenges.  $DS$  pressure, for example, may increase with barriers to entry and information asymmetries being removed to promote market competition, consumers’ rights or stability. At low  $DS$  rate levels, the behavioural reaction is in overall positive, because there is more competition, transparency and legal security, which is good for data generation markets, innovation and investment.

State intervention may solve certain market failures but may also generate new ones if  $DS$  pressure is too high. The other corner solution, ‘ $DS = 100$ ’, represents a ‘Big Brother’ Orwellian world of total data-sharing obligations by all actors in the economy. It implies total negation of data holders’ de facto ownership and individuals’ property and privacy rights. At  $DS = 100$ <sup>8</sup> neither data holders nor citizens have incentives to participate in data generation. Data holders do not find it profitable to invest in innovative services that produce data. Households and firms see their privacy violated and they do not want to pay for digital services and functionalities in data. As a result, there is very small  $DS$  base, and a very low  $DR_{rev}$  and  $K$ . This is analogous to a situation in traditional monetary taxation whereby there would be no officially defined economic activity if the traditional tax rate on such activity were set at 100%. From  $DS = 100$ , reducing data sharing pressure would increase the amount of knowledge generated because it would increase citizens’ willingness to pay in data and data holders’ willingness to invest in digital services. Somewhere between the positive and the negative slopes there is a knowledge-maximising  $DS$  rate,  $DS^*$ . Just as monetary taxation has a tax rate of maximum tax revenue,  $DS^*$  is the  $DS$  rate that maximises  $K$ . If  $DS < DS^*$ , increases in the  $DS$  rate generate a movement along the upwards part of the curve, increasing  $K$ . Up to  $DS^*$  the relationship is positive, while beyond  $DS^*$  the relationship is negative.

The shape of the Data Sharing Laffer Curve captures the conceptualisation of governance as a paradoxical trade-off concept, as described by Chhillar and Aguilera (2022). According to these authors, digital governance is paradoxical because more of one dimension limits another dimension. They identify five governance trade-offs in management literature: governance vs. innovation, reforming vs. strengthening a surveillance state, distributed vs. concentrated power, algorithm efficiency vs. fair data practices and algorithm vs. societal bias. In the case of data-sharing a paradox emerges from the fact that more  $DS$  does not necessarily mean more knowledge. When  $DS$  is below  $DS^*$ , increases in  $DS$  do not limit innovation but encourage it, and help to distribute power and ensure the transparency of algorithms. Beyond  $DS^*$ , data sharing may be counterproductive, may limit innovation and may generate privacy and surveillance concerns.

Data policies can also generate movements of the curve. An economy located below  $DS^*$  means a lack of transparency and underutilisation which negatively influences  $c$  and  $\lambda$ , moving the curve downwards. An economy located beyond  $DS^*$  means high levels of obligatory data sharing where privacy and other citizens’ rights are disregarded. This is akin to a ‘tragedy of commons’ in the data economy. For example, data-sharing may be used to monitor and control citizens’ lives, which erodes the legitimacy of the system itself and negatively influences  $c$  and  $\lambda$ , moving the curve downwards. In both cases the  $IDP = 0$  assumption fails and consumers’ willingness to use their personal data as a means of payment for on-line services falls, thereby reducing the amount of data generated in the economy.

At  $DS = DS^*$ , data policy preserves incentives to invest in data generation, fosters innovation, trust, transparency and the rule of

law and increases confidence in data as a means of payment. This increases *DB* and data-driven innovation, improving technology ( $\lambda$ ). It also facilitates the role of governments, fosters economic stability, reduces market failures and barriers to entry, balances information asymmetries and fosters competition. Carefully designed *DS* policies generate increases in parameter  $c$ , moving the Data Sharing Laffer Curve upwards.

### Does the world fit the semi-circular model?

**The seven assumptions.** Regarding the first and second assumptions, daily activity by households and firms generates a flow of data that data-holders use to produce digital services. This is the case of e-commerce companies such as Amazon, search engines such as Google, social networks such as Facebook, LinkedIn and Instagram, messaging service such as WhatsApp, other services such as Dropbox and Spotify (Kramer and Kalka, 2016, Kumar and Trakru, 2019) and platforms such as Airbnb, Booking, Couchsurfing, Zipcar, Uber, Lyft, BlaBlaCar, TaskRabbit, myTaskAngel, Freelancers, etc. Users generate data when they search, buy, create a user profile indicating their name, occupation, schools attended, when adding other users as ‘friends’, exchanging messages, statuses, pictures, videos, links, ‘likes’ and other social networks and platforms’ reactions together with the other data derived from user activity (paradata, environmental data or footprints) related to their activity. Daily activity is a data factory that produces data about intentions, acts, personal relationships, health, mood, locations, movements, a diverse amount of economic activities, C2C, P2P, B2B, B2C, etc. In addition, more and more devices contain sensors, more activities generate data and there is an increasing capacity to pump zettabytes of unstructured data towards data holders (The Economist, 2017).

Regarding the third assumption, means of payment in many digital services are personal, accompanied by usage data and sometimes a monetary payment (Evans, 2013, Scott Morton et al. 2019, Tett, 2018, Brynjolfsson et al. 2018). Data holders offer a ‘free to use’ digital service that enables people to reduce search and matching costs. The more data the user is willing to generate, the better the search and matching service and the more efficient the marketing and advertising. As the WTO (2018) points out, the digital economy is not about prices but about data and innovation.

The fourth assumption, on consumers’ utility maximisation at  $IDP = 0$ , refers to monetary terms. It does not mean that individuals and companies are giving away their data for nothing. They barter data for the functionality of a digital service. Nor does it imply that individuals are losing monetary value, because most individuals and companies are unable to aggregate and analyse data or create and capture value from them (Wiener et al. 2020). According to Wiener et al. (2020), many companies are increasingly trying to leverage *BD* but there is a ‘deployment gap’ and despite the enormous potential of *BD* across many industries, its actual deployment remains scant.

At the initial stages of digitalisation, this is a realistic assumption. Obar and Oeldorf-Hirsch (2018) study users’ behaviour in reading privacy policies and terms of use. They show that many skip reading and most miss important points (Cakebread, 2017). This shows that most users do not consider their personal data a valuable asset and accept the barter exchange because not participating in some functionalities may often have a high social cost (Bolin and Andersson Schwarz, 2015). Chhillar and Aguilera (2022) refer to the illusion of informed consent with data practices buried in fine print. Not only individuals but also many companies mismanage their intangible assets and conventional accounting systems ignore them (Adams, 2019, Govindarajan et al. 2018). Digitalisation and

*BDBM* evolve overtime (Wiener et al. 2020) and the inability to aggregate, analyse and create and capture value from data may be a focus of future governance.

Data markets corroborate that the value of individual data before aggregation is close to 0. According to the *Financial Times* (Steel, 2013, Steel et al. 2013), data brokers’ pay between EUR 0.0005 and EUR 0.66 (calculations made in October 2018) for data on individuals. Data quality and price depend on the amount of detail: the more observable characteristics and aspects of life they contain, the more valuable data are.

Regarding the fifth assumption, about data holders as profit maximisation companies that consider individual data as a valuable asset, they very often obtain direct monetary compensation not from digital services but from ‘de facto ownership’ of data. They use data to produce knowledge about patterns, regularities and structures of human behaviour and activities (Redman, 2018, Dosis and Sand-Zantman, 2018, Jones and Tonetti, 2018, Scott Morton et al., 2019, Boisot and Canals, 2004, Duch-Brown et al. 2017, Vigo, 2013, Duch-Brown, 2017a, b, c). Individual data are almost valueless in isolation (Steel et al. 2013). Only by having very large pools of data points from, perhaps, hundreds of millions of individuals are data holders able to derive value from those data (Worstell, 2017), and that value is only realised after knowledge extraction. Kumar and Trakru (2019) show that data mining, machine learning and natural language processing are the main AI components which are becoming ever more central to the workings of the global economy.

Regarding network effects, although  $IDP = 0$  and the value of individual data is close to zero, market valuation of data factories is different, especially if they are able to generate network effects and attract more users. This is illustrated by the huge amounts that have been paid for (apparently) non-profitable companies which have developed services with network effects (Bond and Bullock, 2019, Kaminska, 2016, McArdle, 2019). Instagram and WhatsApp’s acquisitions by Facebook in 2012 and 2014, respectively and Google’s acquisition of YouTube in 2006 are good examples. The reduction of competition (via predation of markets) and the ability to generate network effects may therefore explain the valuation of these data factories. Acquisitions are also data-quality driven, as they expand data collection to other individuals and realms of life. Conventional accounting systems largely ignore data but data are the primary drivers of financial performance. Some authors even go so far as to assert that conventional accounting numbers are irrelevant for digital companies<sup>10</sup> (Adams, 2019, Govindarajan et al. 2018) because their principle value creators are increasing returns to scale on *intangible investments and network effects*. New methods of valuing intangible assets show that a strong intangible asset position delivers sustainable competitive advantages such as network effects<sup>11</sup>. For investors, the most important characteristics of digital firms are market leadership and network effects that might lead to a ‘winner-take-all’ structure (Govindarajan, 2018). According to the WTO (2018), the nature of competition in digital markets is materially different from competition in traditional markets as it tends to be based on innovation, where data play a fundamental role, rather than on pricing.

Regarding the seventh assumption, in knowledge production where *BD* and *AI* are used to address as many aspects of life and individuals as possible, diminishing returns to scale may never appear. First, there are very high fixed costs and negligible variable costs (Duch-Brown, 2017a). Fixed costs refer to connectivity infrastructure such as broadband (UNCTAD, 2017), research and development, data centres, cloud computing arms and data refineries to handle data generation, collection and processing (The Economist, 2017). Second, the more data that are fed into self-optimising AI algorithms (Silver et al., 2017), the

more AI improves. Data show decreasing returns to scale when prediction aims for a limited number of variables (Varian, 2013, 2018). But this is not the case where BD and AI methods are increasingly prevalent (Kumar and Trakru, 2019). There is a clear positive impact of volume and variety in data value. Wheeler (2021) defines the 'scale and scope reinforcing loop' as the never-ending process in which data produce new products, which produce new data, thus speeding the pace of change beyond the capacities of the industrial era. Identifying where economies of scale give way to diminishing returns is an empirical issue on which there is little evidence (Codagnone and Martens, 2016). Several studies and reports support the contention of a lack of competition and the existence of scale and scope due to low marginal cost and two-sided network effects, finding evidence of monopolisation or monopoly power (George J S Center, 2019, Competition and Markets Authority UK, 2020, Furman, 2019, Ghosh, 2020, Wheeler, 2021).

Costs of diversification and innovation may oppose scale, scope and concentration in services, products and data production markets. However, they do not seem to oppose concentration in knowledge production using BD and AI because this is a specialisation in itself. For example, Facebook, WhatsApp and Instagram may compete as social networks with different specialisations in digital service markets, but knowledge is extracted more efficiently if the data obtained are merged and analysed using the same tools and methods. The platform economy has its limits (Azzellini et al. 2019) but data holders expand into physical production and sectors where platforms are not yet taking over (Govindarajan, 2018). Amazon's acquisition of Whole Foods, which extends data collection to offline activities, illustrates how knowledge extraction using AI is a specialisation in itself (Hirsch, 2018, Krugman, 2014): it was a data-driven acquisition that expanded Amazon's data collection to offline activities. Amazon increased both X (the type of activities on which it was able to collect data) and N (the type of consumer that it was able to follow). Sofa Sounds' partnership with Uber and AirBnB is another example of a data-driven expansion without high diversification costs because it does not imply a new specialisation. During the process of obtaining European Commission approval to merge Facebook and WhatsApp (European Commission, 2017), Facebook pledged that it would not merge user bases but, as far as we know, no authority has been charged with seeing that it does not do so. Other data-driven acquisitions, interconnections and partnerships between companies resemble a spaghetti bowl and may reflect incentives to centralise knowledge production. For example, MasterCard Advisors are IBM Watson partners. In principle, PayPal is a competitor of MasterCard, but MasterCard owns a percentage of PayPal and PayPal is a Facebook partner. Facebook has received investment from PayPal. In China, the same company integrates social networks and the payment industry through 'WeChat', which, in a single application, offers services like those of Instagram, Facebook and WhatsApp together with payment services. IBM's acquisition of the Weather Company in 2015 illustrates that concentration goes beyond personal data to information on context variables that determine consumer behaviour.

BD and AI reinforce each other and thus enhance the concentration process. Data holders expand investments in companies able to generate data but also in AI companies, with Google's acquisition of world AI leader DeepMind in 2014 being a case in point. DeepMind also has access to public records through its agreement with the United Kingdom's National Health Service (Lomas, 2019). Another example is Facebook's investment in DeepText, an AI natural language processor able to learn the intentions and context of users in 20 languages, and in

face recognition technologies. In general, data are a critical ingredient for feeding AI models and innovation (OECD, 2019). Expansion also affects mobile devices and gadgets such as smart watches that generate more data (Govindarajan, 2018).

In addition, there is evidence of market concentration in the global economy (McKinsey, 2019, 2018, Scott Morton et al. 2019) and the digital sector (OECD, 2019, UNCTAD, 2017). A shrinking number of companies dominates an increasing number of industries. This is accompanied by declining start-up growth and less financial resources for them, fewer young, high-growth firms and growing inequality (Khan, 2017, Porter, 2016, Jarsulic et al. 2016, Decker et al. 2018, The Economist, 2018).

If the fifth, the sixth and the seventh assumptions hold, knowledge generation from BD and AI is a natural monopoly and data holders are in a race towards  $X = \text{everything}$  and  $N = \text{All}$  to gain market power from network effects. If this is the case, there should be signs of creative destruction, price discrimination, market power and investment attraction. Evidence of these phenomena can be considered as empirical clues that underpin the usefulness of our theoretical model and its ability to capture reality.

Data holders' activity generates innovation and expansion of AI, which generates a process of creative destruction (Uber vs. taxis, Airbnb vs. hotels, sharing vs. specialisation, etc.). Creative destruction also affects services traditionally provided by the public sector such as public transport (Evgeny, 2015), health care (Carrie Wong, 2019), banking (Mercola, 2020) and national defence (Brustein and Bergen, 2019).

Access to data generates information asymmetries that open up opportunities for price discrimination, steered consumption and unfair competition in sectors other than knowledge generation (White House, 2015, Ursu, 2018, Mikians et al. 2012, Shiller, 2014, Chen et al. 2015, Möhlmann and Zalmanson, 2017, Uber, 2018, Ezrachi and Stucke, 2016). Discrimination can go beyond prices and lead to unfair treatment and discrimination in general (Isaac, 2017, Wong, 2017). Asymmetric information may also foster predatory pricing and monopsony behaviours (Bensinger, 2012, Bond and Bullock, 2019, Kaminska, 2016, McArdle, 2019, Codagnone and Martens, 2016). Regarding the rule of law, services that emerge in the data economy, especially in the sharing economy, challenge aspects such as consumer protection, professional licences, working conditions, regulations vs. informal supply of services, (Hall and Krueger, 2015, Cook et al. 2018), quality standards (Codagnone and Martens, 2016, Vaughan and Hawksorth, 2014, Malhotra and Van Alstyne, 2014) and tax avoidance (TJ) (D'Andria, 2019). In addition, some hedge funds operating in markets around the world employ a combination of AI models and use BD lakes and human intelligence to obtain privileged information about the economy (Grassegger and Krogerus, 2017, Kosinski et al. 2013, Kee, 2018, Cadwalladr, 2017, Zuboff, 2019). Literature has reported lack of transparency as illustrated by the black box of semi-circular flow of the data economy. According to Chhillar and Aguilera (2022), 'algorithmic decision-making has shown to outperform humans in several activities such as trade efficiency, returns to investment, marketing, fraud detection, credit scoring, weather forecasting, statistical analyses... however, algorithms can also behave in an inaccurate and biased way and are characterised by their opacity and weak accountability.'

As for attracting investment, Khan's anti-trust paradox<sup>12</sup> (Khan, 2017) puts the accent on the long run and the limitation on recognising harm to competition from short-term prices and outputs. By contrast, the long-run competitive advantages of knowledge generation and innovation are an important driving force behind concentration and attracting investment. Data holders and investors maximise data collection and expand their



data collection infrastructures because access to data and knowledge shapes globalisation, innovation and the distribution of wealth and power (OECD, 2019, ITU, 2018, Freeman, 2013, Lehdonvirta, 2022). The data economy is not a small add-on to the circular flow but a key element of long-run growth, market power and dominance (Arthur, 2011). In fact, although dominance has grown also thanks to mergers and proprietary market places, allowing data holders to crush competitors, favour their rankings and sell their own brands, digital giants often have meagre short-term profits but set their priorities on intensive data-hungry (long-run) growth (Facebook, 2014, Statista, 2015, Khan, 2017, Lehdonvirta, 2022).

The semi-circular flow model is also consistent with BDBM literature. In their recent BDBM literature review, Wiener et al. (2020) report a 'deployment gap', which they define as a paradox between two facts: on the one hand the enormous potential of BD across industries and, on the other hand, the observation that actual deployments of BDBM remain scant. Several studies report high degrees of vertical integration in large organisations including data supply, storage, processing and even the smart-devices market. However, many organisations, especially SMEs, remain in a limbo stage, unable to deploy and internalise the capability to use and leverage BD. This resembles the concentration process described above (Chen et al. 2015, and Schroeder, 2016). According to Wiener et al. (2020), traditional organisations struggling to leverage BD coexist with large data-driven companies that outperform them.

**Governments' reactions in terms of Eq. (1).** The literature supporting disequilibria is described in Section 'Disequilibria in a data-intensive economy' (OECD, 2019, Liem and Petropoulos, 2016, UTI, 2018, OECD, 2019, D'Andria, 2019). Data holders are very attractive to investors ( $I\uparrow$ ) and this, together with tax avoidance ( $T\downarrow$ ), generates a disequilibrium such as

$$S + T + M < I + G + X$$

Expanding creative destruction to the state itself: reductions in public expenditure ( $G\downarrow$ ) lead the economy back to an equilibrium where the role of the state diminishes. Following a traditional view of the economy, some countries have approached the issue by trying to increase unilateral taxes ( $T\uparrow$ ) and fines (Pratley, 2018, Sandle, 2018). Countries and international institutions are devising ways of taxing digital activity (European Commission, 2016, OECD, 2019, D'Onfro and Browne, 2018, European Commission, 2018a, 2018b, Khan and Brunnsden, 2018, Gold, 2019) and/or collect money through antitrust fines (European Commission, 2017). In Germany, for example, failure by Facebook to remove banned content within 24 h results in fines of up to 50 million Euros. In a more user-centric approach, Posner and Weyl (2018) propose that agents could be compensated for the data that they generate just as they are compensated for their labour or in the form of a dividend (Ulloa, 2019). Such compensation still has a 'monetary' view of the data economy and does not take into account difficulties in pricing individual data. These reactions increase the financial power of governments ( $T\uparrow$ ) and may help to bring the economy back to equilibrium. However, they do not affect any of the parameters of Eq. (6) that determine knowledge production ( $K$ ).

As an alternative, it has been argued that breaking up companies such as Amazon, Facebook and Google (Alphabet) would generate enhanced competition. Breaking up, however, entails an opportunity cost for the whole of society. It would imply duplication of resources and—potentially—lower innovation. In other words, society would not take full advantage of economies of scale and scope. As pointed out by the anti-trust paradox (Khan, 2017), these traditional views fail to include the

data and knowledge dimension of the economy. Data holders' 'de facto ownership' operates as a 'breastplate': a shell that prevents additional knowledge production. The United Nations (UN, 2014, UTI, 2018) has reported growing inequalities in access to data, information and the ability to use them. Distribution of information generates asymmetries and fosters inequalities (Duch-Brown et al. 2017, Stiglitz, 2001). However, there is no authority or institution in charge of removing barriers to data access to promote efficiency, equity, stability, data capitalism societal challenges and redistribution of data and knowledge.

BDBM literature also supports action in the data dimension of the economy. Wiener et al. (2020) report two overarching findings from their BDBM literature review: weak theoretical underpinnings in current BDBM studies and a strong emphasis on value creation at the expense of value capture, neglecting other stakeholders, in the BD life cycle from data collection to analysis and use. They conclude that data sharing across industries is an extremely substantial opportunity to assess the bright and dark sides of BD and society. Our model helps to bridge the theory gap and supports data sharing to increase knowledge creation and distribute value across the economy as a way of avoiding concentration inefficiencies.

**Where is the economy located in terms of data sharing?** There are three AI leaders worldwide: the USA, China and the EU (European Commission, 2018a), with the EU lagging behind the first two.

In the USA and Europe, corporate data holders decide for what and to whom they give access to data. For example, access to data by the social sciences community is very limited. First, they can explore the surface of the digital economy by web crawling (Pedraza et al. 2019). Second, they can benefit from (non-disclosure) agreements, but such agreements may generate a data divide among scientists, jeopardising replicability and FAIR (Findable, Accessible, Interoperable, Reusable) principles (Wilkinson et al. 2016, Taylor et al. 2014, Codagnone and Martens, 2016, Malhotra and Van Alstyne, 2014, Hall and Krueger, 2015, NSF, 2017). Third, they can use the data crumbs that data holders make available to activate the research community and thus obtain new perspectives on their own business. This is the case of Google trends and other 'data philanthropy' initiatives (Pawelke and Tatevossian, 2013). Internet searches contain insights into diverse human activities (Askitas and Zimmerman, 2009, 2011a, 2011b, 2011c, Choi and Varian, 2011, Askitas, 2015) but the data released are not enough to build and test consistent, stable models (Artola et al. 2015), as shown by Google flu predictions (Ginsberg et al. 2009, Butler, 2013).

Although all these data have given rise to thousands of academic papers, the economy is at or close to *DS rate* = 'de facto ownership', i.e. below  $DS^*$ . The amount of knowledge produced in the economy is that which maximises data holders' profits and their data collection; further knowledge production relies on their good intentions via data philanthropy (Taylor et al. 2014, Einav and Levin, 2013). Current data sharing is not the result of state intervention or individuals' exercise of their data ownership rights. Regulatory authorities and the scientific community remain unable to fully tap into innumerable aspects of digital policy-making, which keeps knowledge about data capitalism's societal challenges and dilemmas underexplored (Khan, 2017, Scott Morton et al. 2019, Taylor et al. 2014, Butler, 2013, Artola et al. 2015, Lazer et al. 2014).

Thus, both the EU and the USA have similar *DS rates*. The amount of knowledge produced by the EU is lower than that of the US but its position is still good in terms of AI publications (European Commission, 2018a). In the EU the GDPR regulation

supports data portability (according to art. 20 of the GDPR, data can be transferred from one controller to another) and the European Commission's Data Strategy supports data sharing (European Commission, 2020). US knowledge production is higher because of the strength of American corporations, which possess huge BD lakes and AI, which has a spillover effect in  $k$  and  $\lambda$ . The US is the world leader in start ups and venture capital.

In China, centralisation with no clear distinction between data holders, the state, supervision and surveillance and the presence of multifaceted tools such as WeChat gives the country a competitive advantage in developing huge BD lakes and economies of scale and scope. China is the world leader in turning research into patents (European Commission, 2018a). China's data fiscal pressure is probably beyond  $DS^*$  and on the downward sloping part of the curve. According to Freedom House (2019), China is the world leader in developing and exporting social media surveillance tools.

None of those three situations is static. As digital literacy evolves, actors in the economy become more aware of implications of payments using data. Consumers may change their data generation behaviour when there is a lack of transparency or when data are used against consumers' interests to support unfair competition, price discrimination, manipulation, political distortion, surveillance, etc. (Toscano, 2019, Lyon, 2014). Lack of transparency reduces trust and thus individuals' willingness to 'pay' in data, the  $IDP = 0$  assumption fails and parameter  $c$  goes down. For example, as of October 2018, 74% of Facebook users were unaware that advertisers were able to make use of their lists of interests for targeting purposes, but after the Cambridge Analytica scandal 54% of users adjusted their privacy settings, started using Facebook less frequently or even left the app (Gramlich, 2019). As data literacy evolves, economic actors will eventually demand legal security to use their data as a means of payment, just as payments in non-reliable currencies are not accepted.

### Discussion: how? what? who?

From the foregoing it can be concluded that an increase in data sharing would increase the knowledge available to the economy and society concerning the societal challenges of data capitalism and its governance. This raises at least three questions.

#### How can governance modalities activate data sharing?

According to Chhillar and Aguilera (2022) there are four possible governance modalities for tackling the societal challenges posed by data capitalism: norms, law, market and architecture.

Regarding social norms, digitalisation of the economy is a dynamic process, so social norms, stigma and sensitivity towards societal challenges are also dynamic. Societies have different levels of tolerance towards issues such as privacy, inequalities and innovation and may support different levels of data sharing accordingly. According to Zuboff (2019), as digitalisation evolves, so the power of data holders to infer opinion and actions increases (Connolly 2016). In that case, we can expect very little criticism towards data capitalism and little support for data sharing. Data holders, as de facto owners of data, will continue sharing their data sets as part of their philanthropic data marketing just as they are already doing. If more criticism emerges, it will be in data holders' own interest to generate a degree of further transparency and trust in the data generation process.

The 'law' modality refers to mandates from public institutions, which are enforced by an authority. The state can promote enhanced data flows via incentives and disincentives linked to a data tax system as it does with traditional taxation. So far this

modality has focused very much on privacy, with the EU's GDPR being the main example. Governments, in the exercise of their monopoly on power, can increase data sharing pressure depending on different political views of equity, efficiency, stability, privacy, transparency and surveillance applied to the societal challenges posed by data capitalism. For example, governments with a stronger emphasis on equity may incorporate data sharing into their competition laws in order to narrow the deployment gap, while governments focused on efficiency may be more protective with economies of scale and scope, allowing bigger data lakes. Similarly, governments with an emphasis on stability may promote data sharing so as to improve knowledge on forecasting models and analyses of the economic cycle.

The 'market' modality refers to supply and demand. In competitive markets consumers can choose the companies that best cover their preferences in terms of societal issues such as privacy or equality. As reported above, in the digital economy market concentration and lack of competition enable data holders to operate on 'take it or leave it' terms.

The idea of individuals being able to decide on terms and conditions goes beyond the market into governance architecture. Data sharing based on consumers' individual decisions needs all four modalities to operate together. From an economic intuition point of view, when consumers maximise their utility beyond the  $IDP = 0$  assumption and data holders' interests, higher levels of trust ( $c$ ) should result (Jones and Tonetti, 2018). This would move the Laffer Curve upwards. From a legal point of view, clear and real ownership rights have always been a prerequisite for a well-functioning market economy and for maximisation of consumers' utility. As owners of their personal data, citizens should be empowered and encouraged to decide who should be given access to their data. The EU's GDPR (European Union, 2016), which came into force in 2018, is an overall legal benchmark that sets the legal basis for a user-centric approach<sup>13</sup>. In fact, the GDPR seeks to facilitate the free flow of personal data with the goal of protecting the rights of citizens. According to De Hert et al. (2018), the right to data portability is a novel feature of the GDPR that forms the basis for additional regulation beyond data protection and towards competition law or consumer protection. In practice, free movement of data and data portability are very limited: users are the legal owners (European Union, 2016, Jones and Tonetti, 2018) but data holders collect, control and draw value from their data. Implementation and full exercise of portability rights may require higher levels of data literacy among citizens and an architecture. However, this does not seem to be incorporated into social norms. Regarding enabling tools and empowerment of informed decisions by citizens<sup>14</sup>, one possible architecture could be PDSs. PDSs are an emerging business model that seeks to facilitate users' exercise of their personal data ownership rights and give users more options to control their data in terms of permissions to access them and generation of value (Bolychevsky and Worthington, 2018).

**What would efficient data sharing policy look like?** A Pareto-efficient data sharing policy would improve the situation of the actors who are the beneficiaries of interventions, mainly households and firms, and would have positive externalities for society as a whole without generating negative consequences on efficient resource allocation or discouraging investment and R&D activities. A Pareto-efficient intervention does not rival data holders' activities. Traditionally, governments impose community standards and support consumption of merit goods such as education, and ban or discourage demerit ones such as illegal drugs and tobacco, respectively. The idea behind merit and demerit goods is that a well-informed society is in a better position to identify the

amount needed of certain goods that have positive externalities for societal well-being, citizens' safety and economic growth (Lucas, 1988, Munich and Psacharopoulos, 2018). If data-sharing policy follows a similar rationale, it should promote consumers' and citizens' ability to make informed decisions by generating and disseminating more knowledge about the digital economy black box.

Determining what is merit and what is demerit knowledge lies outside the scope of this paper. It is a subjective discussion that has varied over time and across cultures. Using economic welfare as a measure and the perspective of western market democracies, the distinction might be relatively straightforward. Knowledge would be considered to lead to demerit outcomes when used to violate privacy, generate market power and set barriers to entry, generate information asymmetries or unacceptable distribution of wealth, control market places and damage competition, charge unfair fees or prices, monitor and control citizens' lives, manipulate political campaigns or impose excessive regulations limiting innovation. By contrast, knowledge could be considered to lead to merit outcomes when used to innovate and reduce market frictions, information costs and asymmetries, generate better matches between supply and demand and facilitate the full utilisation of private assets that would otherwise be idle. Knowledge is also a merit good if used to study market structures and anti-trust concerns, adapt the existing legal corpus to the new digital reality, find ways to foster competition, promote transparency, the rule of law and enforcement, forecast economic cycles and deliver nimbler and faster anti-cyclical policies. Light could be shed on data capitalism governance challenges (Chhillar and Aguilera, 2022) by merit knowledge.

There are many grey areas where boundaries may be blurred and not so clear-cut: consider scientific research that might require barriers to entry for some time, as is recognised in the IP literature (Ilie, 2014). Or consider the context of COVID-19, where using mobile apps to track the virus may have implied some invasion of privacy (Zhang et al. 2020).

**Who should have access to data?** As Varian (2018) points out, data access is probably more relevant than data ownership. Even if individuals' data ownership rights and data portability are recognised and PDSs are promoted and facilitated, individuals in general have no ability to extract knowledge from data. There are many public entities (potentially) able to produce merit and non-rival knowledge if they are given access to data, e.g. central banks, antitrust authorities, labour inspectors, the scientific community and other agents that are not direct competitors of data holders but who could contribute to a better-informed society and move the  $DS$  and  $K$  along the positive slope of the curve and increase  $c$ . Central banks could improve our understanding of the economic cycle. Antitrust authorities could enhance research on sources of unfair competition, deliver antitrust policies and balance information asymmetries specific to a data intensive economy. Labour inspectors could study working conditions in the platform economy. The scientific community has shown that, if it is not limited by data access, it can enhance knowledge about many research topics and phenomena (Schroeder and Cows, 2014) and find ways to tackle societal challenges. In addition, as in any other market, promoting competition and removing barriers to entry increases efficiency, in part by encouraging entry into the market via new start-up enterprises.

Some authors have proposed the establishment of an international data authority (Martens, 2016, Scott Morton et al. 2019, Askitas, 2018) or an international digital alliance (Wheeler, 2021). Some of the issues described above are conducive to an international institution rather than individual national efforts.

Primarily, because of the supranational nature of data holders. Secondly, such a data authority would need to be flexible and to embrace techniques that mirror flexible management by data holders themselves. It would operate along lines similar to those described above regarding high fixed costs and economies of scale and scope. Regarding the former, if a data authority seeks to enforce existing data protection and other rights, empower users and protect supply chains and competition while dealing with changing technology and market places, it needs continual investment in R&D. International cooperation can avoid duplication in R&D and investment efforts. Regarding the latter, the amount of data that needs to be analysed and the complexities and limitations of current enforcement tools call for an institution staffed by specialists and data analysts, with an infrastructure able to benefit from scale and scope. Third, such an authority/alliance should be developed in collaboration with data holders rather than in opposition to them. Conversations, negotiations and agreements will be more efficient with a single international organisation as the sole interlocutor than with individual countries or at local level (Scott and Young, 2018, Barzic et al. 2018). Giving democratic legitimacy to such authority will be an additional challenge.

Table 1 shows some examples of potential merit users whose access to data would create knowledge and move the economy along the curve towards  $DS^*$  and, as a consequence, move the data sharing Laffer curve upwards.

## Conclusions

In this paper, we describe a data-intensive economy as an economy where households and firms generate BD in their daily activity, data holders use AI to extract knowledge and services from BD, data are an implicit means of payment with no explicit price formation process, consumers maximise their utility assuming that the value of their personal data is effectively zero, data holders consider data a valuable asset, data generation markets show direct and indirect network effects and knowledge generation using AI and BD shows economies of scale and scope.

In such an economy, knowledge generation using BD and AI shows natural monopoly characteristics leading to concentration, the attraction of investment, creative destruction and price discrimination.

The activity of data holders is supranational and generates different types of macro-economic disequilibria that emerge from difficulties in taxing them, attraction of investment and concentration of activity in a few countries. Governments' reactions focus on traditional disequilibria but do not tackle the data/knowledge/innovation dimension of the economy. Therefore, we continue our theoretical argumentation exploring what a data policy could look like.

According to conventional theory, equilibrium production in monopolistic markets is lower than in competitive ones, so we build upon the idea that government intervention could consist of addressing data leakages and injecting knowledge so as to increase the amount of knowledge available to the economy and to society at large. This idea is in line with the literature that considers that there are governance dilemmas and societal challenges in the data economy that need to be tackled. We specify and draw a Data Sharing Laffer Curve that resembles functions used to analyse other types of intervention, such as corporate taxes or trade tariffs. We assume that the amount of data shared does not affect the amount of knowledge produced in a linear manner. At low  $DS$  rates the relationship is positive and increases in  $DS$  rate increase  $K$ . At high  $DS$  rates the relationship has a negative slope. There is a data sharing level that maximises knowledge production. This idea is in line with governance paradoxes reported in the relevant literature.

**Table 1 Themes, societal challenges (Chhillar and Aguilera, 2022) and data sharing: Merit users ('Who') and knowledge ('What').**

Theme	Societal challenge	Data Sharing: Who and What		
		Merit User	Knowledge creation	
Data capitalism and privacy risk	Power imbalance	Non-rival SMEs and start ups Research/academic community and universities Anti-trust authorities	Innovative functionalities  Market structure measures Competition laws	<ul style="list-style-type: none"> <li>• Narrower deployment gap</li> <li>• Redistribution of power</li> <li>• Innovation</li> </ul>
	Privacy	Privacy-protecting courts	Transparency on data holders' respect for privacy	<ul style="list-style-type: none"> <li>• Antitrust policies</li> <li>• Redistribution of information asymmetries</li> <li>• More competitive markets</li> <li>• Black box transparency</li> <li>• Trust</li> <li>• Algorithmic transparency</li> </ul>
Bias and opacity	Algorithmic bias	Human- AI algorithmic auditors	Technical processes	<ul style="list-style-type: none"> <li>• Trust</li> <li>• Algorithmic transparency</li> </ul>
	Opacity of internal processing	Human- AI algorithmic auditors	Technical processes	<ul style="list-style-type: none"> <li>• Trust</li> <li>• Algorithmic transparency</li> </ul>
Human trust	AI aversion	PDS developers Consumers' associations	Innovative functionalities for consumers' utility maximisation	<ul style="list-style-type: none"> <li>• Trust and digital literacy</li> <li>• More competitive markets</li> <li>• Redistribution of power</li> <li>• Consumer surplus</li> </ul>
	Power asymmetries and algorithmic agency	Labour inspectors	Working conditions	<ul style="list-style-type: none"> <li>• Trust</li> <li>• Algorithmic transparency</li> </ul>
Organisational algorithmic decision making	New occupational responsibilities	Research/academic community and universities	Education programmes	<ul style="list-style-type: none"> <li>• Redistribution</li> <li>• Trust and digital literacy</li> <li>• More competitive markets</li> <li>• Redistribution of power</li> <li>• Consumers' surplus</li> <li>• Critical thinking</li> <li>• New theories</li> </ul>
Economic value creation	Imbalanced property rights division	Central banks Research community SMEs and Start ups	Innovative functionalities	<ul style="list-style-type: none"> <li>• Economic forecasting</li> <li>• New theories</li> <li>• Innovation</li> <li>• Redistribution of power</li> </ul>

We then explore how closely the theory matches the world around us. We report references that support the seven assumptions, concentration, investment attraction, creative destruction and price discrimination. We also conclude that the current level of data sharing in the EU and US is in all likelihood below that which would maximise knowledge. Our main conclusion is therefore that an increase in data sharing would increase the knowledge available to the economy and to society. The knowledge that data sharing could generate would help provide a better understanding of the data economy and help address the societal challenges associated with it in a meaningful way.

The digital economy is a complex reality that affects almost every aspect of human life. This makes considered discussion about its governance difficult. The goal of this paper is to draw up a simple theoretical framework that helps identify and generalise the main market failures, societal challenges and dilemmas of the digital economy so as to facilitate further discussions on its governance in both academic and policy fora. Our theory derives from macroeconomics and is able to capture the main components of data capitalism, incorporating many of the existing references and literature reviews. It incorporates aspects of management research, business and society. It helps to bridge the theory gap and supports the idea of data sharing to increase knowledge creation and distribute value across the economy as a

way of avoiding concentration inefficiencies and a lack of clear governance. This paper gives both policy-makers and researchers strong arguments for claiming stronger access to data, and focuses on the enormous possibilities that data sharing can open up for research and the opportunity cost of not doing so.

Our theory also has several limitations and caveats that we leave open for further discussion. First, we do not claim that the references provided to support our assumptions are conclusive. Rather, we hope to stimulate further research which could further investigate the main implication that emerges from our paper, i.e. the goal of increasing the current data sharing level in the economy. Second, the Discussion section leaves three issues open: the most efficient ways to generate optimal levels of data 'leakage', what kind of knowledge should be generated and who should be in charge of it. These discussions will be the subject of our future research.

Received: 4 November 2022; Accepted: 6 September 2023; Published online: 23 September 2023

**Notes**

<sup>1</sup> Digital giants are also geographically concentrated and serve markets through trade. This challenges many countries with growing imports (M1), exacerbating existing

- current account imbalances. These imbalances are beyond the scope of this paper and, in any case, still do not cover the data/knowledge/innovation dimension of the economy that we define.
- 2 The semi-circular flow of the economy does not explicitly include data generated by firms and citizens in their interaction with the public sector. Cases in point are social security records, administrative data and medical records at public health services. Data holders have a competitive advantage in using AI technologies to obtain knowledge extraction technologies. They very often successfully compete in public tenders and have access to public-sector data (Lomas, 2019).
  - 3 The semi-circular flow of the economy does not explicitly include data generated by firms and citizens in their interaction with the public sector. Cases in point are social security records, administrative data and medical records at public health services. Data holders have a competitive advantage in using AI technologies to obtain knowledge extraction technologies. They very often successfully compete in public tenders and have access to public-sector data (Lomas, 2019).
  - 4 This does not necessarily mean that collusion will occur, or that it is necessarily legal or socially desirable. It simply means that there are rents available to the firms present under oligopoly if collusion can be 'achieved'.
  - 5 Generating a direct flow of data back to firms and families would not solve information asymmetries because, in general, they do not have the ability to extract knowledge from BD.
  - 6 We follow literature studying corporate tax revenues, see Kimberly (2007).
  - 7 Similar graphs and functional forms are used to study other types of intervention such as trade tariffs. For example, see the representation of the relationship between well-being and trade tariffs in Krugman and Obstfeld (2006), page 227. At the optimal tariff the marginal gain from an improvement in the terms of trade is equal to the loss of efficiency derived from the distortion in production and consumption.
  - 8 At  $DS = 0$  and  $DS = 100$ , knowledge generated is  $>0$  because BD and AI are not the only sources of knowledge and some degree of knowledge production,  $k$ , exists regardless the amount of data generated and the level of data sharing.
  - 9 Data brokers specialise in collecting personal data and data on companies and selling them on to third parties.
  - 10 Govindarajan et al. 2018 report that 'Uber's value is estimated between \$48 and \$70 billion, despite reporting losses over the last two years. Twitter reported a loss of \$79 million before its IPO, yet it commanded a valuation of \$24 billion on its IPO date in 2013. For the next 4 years, it continued to report losses. Similarly, Microsoft paid \$26 billion for loss-making LinkedIn in 2016, and Facebook paid \$19 billion for WhatsApp in 2014 when it had no revenues or profits. In contrast, industrial giant GE's stock price has declined by 44% over the last year, as news emerged about its first losses in the last 50 years'.
  - 11 Adam (2019) uses Instagram's sale to Facebook as an example of intangible asset value and illustrates the difference between account value and driving value, taking into account intangible assets such as data: 'Instagram was 20 months old, had no revenue, effectively no assets and 12 employees. A traditional cost or cash flow basis method would have said this company was worth \$0 yet it was sold to Facebook for \$1 billion. (...) Today Instagram would be worth more than \$100 billion'.
  - 12 The (Amazon) anti-trust paradox refers to a situation arising from a strategy used by digital giants in which a combination of staggering growth and wide expansion, meagre profits and below-cost pricing is used by companies to position themselves at the centre of e-commerce and as essential infrastructures upon which other businesses depend. The welfare framework and anti-trust regulations fail to recognise the resulting anticompetitive market dominance. This occurs in a context where investors reward the pursuit of growth versus profits, enabling predatory pricing. By controlling the essential infrastructure, digital giants also collect and exploit the data generated on it.
  - 13 GDPR Art 51 on empowering users.
  - 14 According to the GDPR (European Union, 2016), lawful processing (Art 6) of data can be based, for example, on consent (Art 7), which has to be given for each specific, explicit purpose (Art 5.b). Data holders who have received informed consent from a data subject can only use the data for the specific, explicit purpose for which consent is given.
- ## References
- Acemoglu D, Robinson J (2012) *Why nations fail: the origins of power, prosperity, and poverty*. New York: Crown Bus
- Adams P (2019) The missing trillions: valuing intangible assets. EverEdge Global The missing trillions: valuing intangible assets | EverEdge Global
- Arthur WB (2011) 'The second economy' *Mackinsey Quarterly*, October 2011 [http://www.mckinsey.com/insights/strategy/the\\_second\\_economy](http://www.mckinsey.com/insights/strategy/the_second_economy)
- Artola C, Pinto F, Pedraza Pde (2015) Can internet searches forecast tourism inflows? *Int J Manpower* 36(1):103–116. <http://www.emeraldinsight.com/doi/pdfplus/10.1108/IJM-12-2014-0259>
- Askitas N (2015) Google search activity data and breaking trends. *IZA World Labor* 2015:206. <https://doi.org/10.15185/izawol.206>
- Askitas N, Zimmermann KF (2009) Google econometrics and unemployment forecasting. *Appl Econ Q* 55(2):107–120. <https://doi.org/10.3790/aeq.55.2.107>
- Askitas N, Zimmermann KF (2011c) Nowcasting business cycles using toll data. *J Forecast* 32(4):299–306. <https://doi.org/10.1002/for.1262>
- Askitas N (2018) A data tax for digital economy. *IZA World Labour* <https://wol.iza.org/opinions/a-data-tax-for-a-digital-economy>
- Askitas N, Zimmermann KF (2011a) Health and well-being in the crisis. *IZA Discussion Paper* no 5601, March 2011, Institute of Labor Economics
- Askitas N, Zimmermann KF (2011b) Detecting mortgage delinquencies. *IZA Discussion Paper* no 5895, July 2011, Institute of Labor Economics
- Azzellini D, Greer I, Umney C (2019) Limits of the platform economy: digitalization and marketization in live music. Working Paper *Forschungsförderung*, number 154, August 2019, Hans Blockler Stiftung
- Barzic G, Rose M, Rosemain M (2018) French officials are going to work at Facebook for 6 months *World Economic Forum*. <https://www.weforum.org/agenda/2018/11/france-to-embed-regulators-at-facebook-to-combat-hate-speech/>
- Bensinger G (2012) In Kozmo.com's failure, lessons for same-day delivery. *Wall Street J*, <https://blogs.wsj.com/digits/2012/12/03/in-kozmo-coms-failure-lessons-for-same-day-delivery>
- Board S, Lu J (2018) Competitive information disclosure in search markets. *J Political Econ* 126, 5. <https://doi.org/10.1086/699211>
- Boisot M, Canals A (2004) Data, information and knowledge: have we got it right? *J Evol Econ* 14(1):43–67. <https://doi.org/10.1007/s00191-003-0181-9>
- Bolin G, Andersson Schwarz J (2015) Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data Soc*, 2(2). <https://doi.org/10.1177/2053951715608406>
- Bolychevsky I, Worthington S (2018) Are personal data Stores about to become the next big thing? <https://medium.com/@shevski/are-personal-data-stores-about-to-become-the-next-big-thing-b767295ed842>
- Bond S, Bullock N (2019) Uber IPO prospectus shows ride-hailing revenues stalled. *Finan Times*, 11 April <https://www.ft.com/content/c68d3662-5c76-11e9-939a-341f5ada9d40>
- Brustein J, Bergen M (2019) Google's defence dilemma: the company wants the military business, most of its employees don't. *Bloomberg Businessweek*, 25 Nov 2019, Europe edition, pp 38–43
- Brynjolfsson E, Eggers F, Gannamaneni A (2018) Using massive online choice experiments to measure changes in well-being. *NBER Working Paper* 24514 <http://www.nber.org/papers/w24514>
- Butler D (2013) When Google got flu wrong. *Nature* 494(7436):155. [https://www.nature.com/news/polopoly\\_fs/1.12413!/menu/main/topColumns/topLeftColumn/pdf/494155a.pdf](https://www.nature.com/news/polopoly_fs/1.12413!/menu/main/topColumns/topLeftColumn/pdf/494155a.pdf)
- Cadwalladr (2017) Robert Mercer: the big data billionaire waging war on mainstream media. *The Guardian*, 26 Feb 2017. <https://www.theguardian.com/politics/2017/feb/26/robert-mercer-breitbart-war-on-media-steve-bannon-donald-trump-nigel-farage>
- Cakebread C (2017) You are not alone, no one reads terms and conditions of service agreements. *Business Insider*, 15 Nov 2017. <https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11>
- Carrie Wong J (2019) Will Google get away with grabbing 50m Americans' health records? *The Guardian*, 14 Nov 2019. <https://www.theguardian.com/technology/2019/nov/14/google-healthcare-data-ascension>
- Chen L, Mislove A, Wilson C (2015) Peeking beneath the hood of Uber. <https://doi.org/10.1145/2815675.2815681>
- Chhillar D, Aguilera RV (2022) An eye for artificial intelligence: insights into the governance of artificial intelligence and vision for future research. *Bus Soc* 61(5):1197–1241. <https://doi.org/10.1177/00076503221080959>
- Choi H, Varian HV (2011) Predicting the present with Google trends. *Econ Record* 88:2–9
- Codagnone C, Martens B (2016) Scoping the sharing economy: origins, definitions, impact, and regulatory issues, Institute for Prospective Technological Studies, Digital Economy Working Paper 2016/0, Ispra, Italy
- Competition and Markets Authority United Kingdom (2020) Online Platforms and Digital Advertising. [https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final\\_report\\_Digital\\_ALT\\_TEXT.pdf](https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ALT_TEXT.pdf)
- Connolly K (2016) Angela Merkel: Internet search Engines are distorting perception. *The Guardian*, 26 October 2016 <https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception>
- Cook C, Diamond R, Hall J, List JA, Oyer P (2018) The gender earnings gap in the Gig economy: evidence from over a million rideshare Drivers. 7 June 2018, Stanford Business School Working Paper no 3637, <https://www.gsb.stanford.edu/faculty-research/working-papers/gender-earnings-gap-gig-economy-evidence-over-million-rideshare>
- Crémer J, Yves-Alexandre M, Schweitzer H (2019) Competition policy for the digital era—Publications Office of the EU (europa.eu), Publication Office of the European Union, ISBN 978-92-76-01946-6 <https://doi.org/10.2763/407537> <https://op.europa.eu/en/publication-detail/-/publication/21dc175c-7b76-11e9-9f05-01aa75ed71a1/language-en>

- D'Andria D (2019) The unbearable intangibility of the internet: taxing companies in the digital era. JRC Science for Policy Brief. [https://ec.europa.eu/jrc/sites/jrcsh/files/fairness\\_pb2019\\_wave03\\_digitaltaxation\\_jrc\\_b2.pdf](https://ec.europa.eu/jrc/sites/jrcsh/files/fairness_pb2019_wave03_digitaltaxation_jrc_b2.pdf)
- D'Onfro J, Browne R (2018) EU fines Google \$5 billion over Android antitrust abuse. CNBC <https://www.cnbc.com/2018/07/10/eu-hits-alphabet-google-with-android-antitrust-fine.html>
- De Hert P, Papakonstantinou V, Maglieri G, Beslay L, Sanchez I (2018) The right to data portability in the GDPR: towards user-centric interoperability of digital services. *Comput Law Secur Rev* 193–203
- Decker RA, Haltiwanger J, Jarmin RS, Miranda J (2018) Where has all the skewness gone? The decline in high-growth (young) firms in the USA, NBER Working Paper 21776 <http://www.nber.org/papers/w21776>
- Dosis A, Sand-Zantman W (2018) The ownership of data. [https://editorialexpress.com/cgi-bin/conference/download.cgi?db\\_name=IIOC2019&paper\\_id=433](https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIOC2019&paper_id=433)
- Duch-Brown N (2017a) The competitive landscape of online platforms. JRC Digital Economy Working Paper 2017-04
- Duch-Brown N (2017b) Quality discrimination in online multi-sided markets. JRC Digital Economy Working Paper 2017-06
- Duch-Brown N (2017c) Platforms to business relations in online platform ecosystems. JRC Digital Economy Working Paper 2017-07
- Duch-Brown N, Martens B, Mueller-Langer F (2017) The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01
- Einav L, Levin JD (2013) The data revolution and economic analysis. NBER Working Paper No 19035, May 2013. <https://www.nber.org/papers/w19035.pdf>
- European Commission (2016) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions — Online Platforms and the digital Single Market Opportunities and Challenges for Europe (COM (2016), 288 final, Brussels, 25.5.2016)
- European Commission (2017) Mergers: Commission fines Facebook €110 million for providing misleading information about WhatsApp takeover. European Commission press release. [http://europa.eu/rapid/press-release\\_IP-17-1369\\_en.htm](http://europa.eu/rapid/press-release_IP-17-1369_en.htm)
- European Commission (2018a) Artificial intelligence: a European perspective. <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective>
- European Commission (2018b) Proposal for a Council Directive laying down rules relating to the corporate taxation of a significant digital presence (COM (2018) 147 final, Brussels, 21.3.2018)
- European Commission (2020) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European strategy for data. Brussels, 19.2.2020 COM(2020) 66 final [https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf)
- European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016, p 1–88)
- Evans DS (2013) Economics of vertical restraints for multi-sided platforms. University of Chicago Institute for Law & Economics. Olin Research Paper No 626. <http://ssrn.com/abstract=2195778>
- Evgeny M (2015) Our cities shouldn't rely on Uber to divide new transport choices. *The Guardian*. 1 Feb 2015 <https://www.theguardian.com/commentisfree/2015/feb/01/cities-need-to-fight-uber-transport-choice-evgeny-morozov>
- Ezrachi A, Stucke ME (2016) The rise of behavioural discrimination. *Oxford Legal Studies Research Paper No 54/2016*. University of Tennessee Legal Studies Research Paper. <https://doi.org/10.2139/ssrn.2830206>
- Facebook (2014) Annual report 2014 [http://www.annualreports.com/HostedData/AnnualReportArchive/f/NASDAQ\\_FB\\_2014.pdf](http://www.annualreports.com/HostedData/AnnualReportArchive/f/NASDAQ_FB_2014.pdf)
- Freedom House (2019) Freedom on the Net 2019: The crisis in Social Media. [https://www.freedomofthenet.org/sites/default/files/2019-11/11042019\\_Report\\_FH\\_FOTN\\_2019\\_final\\_Public\\_Download.pdf](https://www.freedomofthenet.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf)
- Freeman RB (2013) One ring to rule the all? Globalization of knowledge and knowledge creation. NBER Working Papers No. 19301 <https://www.nber.org/papers/w19301>
- Furman J (2019) Unlocking digital competition, report of the digital competition expert panel. March 2019. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/785547/unlocking\\_digital\\_competition\\_furman\\_review\\_web.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf)
- Galenson DW (2017) Economic History. *J Political Econ* 125(6):1747–1752. <https://doi.org/10.1086/694624>
- George JS. Centre for the Study of the Economy and the State (2019). Market Structure and Antitrust Subcommittee Report 21. 1 Jul 2019. The University of Chicago Booth School of Business Committee for the Study of Digital Platforms. <https://research.chicagobooth.edu/-/media/research/stigler/pdfs/market-structure---report-as-of-15-may-2019.pdf>
- Ghosh D (2020) Terms of disservice: how silicon valley is destructive by design. Brookings Institution Press. 16 Jun 2020. <https://www.brookings.edu/book/terms-of-disservice/>
- Ginsberg G, Mohebbi M, Patel R, Brammer L, Smolinski M, Brilliant L (2009) Detecting influenza epidemics using search engine query data. *Nature* 457:1012–1014. <https://doi.org/10.1038/nature07634>
- Gold M (2019) Taxing times. *The Economist*. 28 Oct 2019. <https://eiuerspectives.economist.com/technology-innovation/taxing-times>
- Govindarajan V (2018) Can any one stop amazon from winning the Industrial Internet? *Harvard Bus Rev*. <https://hbr.org/2018/02/can-anyone-stop-amazon-from-winning-the-industrial-internet>
- Govindarajan V, Rajgopal S, Srivastava A (2018) Why financial statements don't work for digital companies. *Harvard Bus Rev*. <https://hbr.org/2018/02/why-financial-statements-dont-work-for-digital-companies>
- Gramlich J (2019) 10 facts about Americans and Facebook. Pew Research Center. Factank News in numbers. 16 May 2019. <https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook/>
- Grassegger H, Krogerus M (2017) The data that turned the world upside down: how Cambridge Analytica used your Facebook data to help the Donald Trump campaign in the 2016 election. [https://www.vice.com/en\\_us/article/mg9vvn/how-our-likes-helped-trump-win](https://www.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win)
- Hall J, Krueger A (2015) An analysis of the labor market for Uber's driver-partners in the United States. Princeton University Working Paper 587 <http://dataspace.princeton.edu/jspui/bitstream/88435/dsp010z708z67d/5/587.pdf>
- Hirsch L (2018) A year after Amazon announced its acquisition of Whole Foods, here's where we stand. CNBC, 15 June, <https://www.cnbc.com/2018/06/15/a-year-after-amazon-announced-whole-foods-deal-heres-where-we-stand.html>
- Ilie L (2014) Intellectual property rights: an economic approach. *procedia economics and finance*. 16, 548–552. [https://doi.org/10.1016/S2212-5671\(14\)00837-5](https://doi.org/10.1016/S2212-5671(14)00837-5)
- Isaac M (2017) How Uber deceives the authorities worldwide. *New York Times*. 3 March 2017 <https://www.nytimes.com/2017/03/03/technology/uber-greyball-program-evade-authorities.html>
- ITU (2018) Assessing the impact of artificial intelligence. International Telecommunications Union, ITU Trends Emerging trends in ICT, Issue Paper No 1, September 2018. <http://handle.itu.int/11.1002/pub/81202956-en>
- Jarsulic M, Gurwitz E, Banh K, Green A (2016) Reviving antitrust. *Ctr. For AM. Prog*. 29 June 2016. <http://www.americanprogress.org/issues/economy/report/2016/06/29/140613/reviving-antitrust>
- Jones CI, Tonetti C (2018) Nonrivalry and the economics of data. Version 0.6, <https://www.gsb.stanford.edu/faculty-research/working-papers/nonrivalry-economics-data>
- Kaminska I (2016) The taxi unicorn's new clothes. *Financial Times*. 1 December 2016. <https://ftalphaville.ft.com/2016/12/01/2180647/the-taxi-unicorns-new-clothes>
- Kee TH (2018) Trump has trained stock market investors. *Market Watch*. <https://www.marketwatch.com/story/trump-has-trained-stock-market-investors-2018-07-20>
- Khan LM (2017) Amazon's antitrust paradox. *Yale Law J* 126(3):710–805
- Khan M, Brunsden J (2018) France and Germany abandon plans for EU digital tax. *Financial Times*. 4 December 2018. <https://www.ft.com/content/fc7330d4-f730-11e8-af46-2022a0b02a6c>
- Kimberly AC (2007) Corporate tax revenues in OECD countries. *Int Tax Public Financ* 14:115–133. 2007
- Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behaviour. *Proc Natl Acad Sci USA* 110(15):5802–5805. <http://www.pnas.org/content/early/2013/03/06/1218772110.full.pdf+html>
- Kramer A, Kalka R (2016) How digital disruption changes pricing strategies and price models. In Khare A, Stewart B, Schatz R (eds). *Phantom ex machina: digital disruption's role in business model transformation*, Springer, Dordrecht. <https://doi.org/10.1007/978-3-319-44468-0>
- Krugman P (2014) Amazon's monopsony is not O.K. *The New York Times*. 19 October 2014. <https://www.nytimes.com/2014/10/20/opinion/paul-krugman-amazons-monopsony-is-not-ok.html>
- Krugman PR, Obstfeld M (2006) *Theory and Policy*. 7th edition. Pearson Education ISBN 10: 84-7829-080-X, ISBN 13: 978-84-7829-101-4
- Kumar T, Trakru M (2019) The Colossal Impact of Artificial Intelligence in e-commerce: statistics and facts. *Int Res J Eng Technol*. 6, 5. <https://www.irjet.net/archives/V6/i5/IRJET-V6I5116.pdf>
- Laney D (2001) 3D data management controlling data volume, velocity, and variety, META Group <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Laney D (2012) Deja Vu: others claiming Gartner's construct for big data <http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-bigdata/>

- Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google flu: traps in big data analysis. *Science*, Vol. 343, 14 March
- Lehdonvirta V (2022) *Cloud Empires: How Digital platforms are overtaking the state and how we can regain control*. Cambridge Massachusetts: MIT Press. ISBN 9780262047227
- Liem C, Petropoulos G (2016) The economic value of personal data for online platforms, firms and consumers. *Bruegel blogspot*. 14 January 2016. <http://bruegel.org/2016/01/the-economic-value-of-personal-data-for-online-platforms-firms-and-consumers/>
- Lomas N (2019) Google completes controversial takeover of Deepmind Health. *TechCrunch*. 19 Sep 2019. <https://techcrunch.com/2019/09/19/google-completes-controversial-takeover-of-deepmind-health/>
- Lucas RE (1988) On the mechanics of economic development. *J Monet Econ* 22:3–42
- Lyon D (2014) Surveillance, snowden, and big data: capacities, consequences, critique. *Big Data Soc*, July–December 2014:1–13
- Malhotra A, Van Alstyne M (2014) The dark side of the sharing economy ... and how to lighten it. *Commun ACM* 57:24–27
- Martens B (2016) An economic policy perspective on online platforms. Institute for Prospective Technical Studies Digital Economy Working Paper 2016/05
- McArdle M (2019) Uber and Lyft are losing money. At some point, we'll pay for it. *Washington Post*. 5 March 2019. [https://www.washingtonpost.com/opinions/uber-and-lyft-are-losing-money-at-some-point-well-pay-for-it/2019/03/05/addd607c-3f95-11e9-a0d3-1210e58a94cf\\_story.html](https://www.washingtonpost.com/opinions/uber-and-lyft-are-losing-money-at-some-point-well-pay-for-it/2019/03/05/addd607c-3f95-11e9-a0d3-1210e58a94cf_story.html)
- Mckinsey (2018) Superstars. The dynamics of firms, sectors, and cities leading the global economy. McKinsey Global Institute Discussion Paper. [https://www.mckinsey.com/~/media/mckinsey/featured%20insights/innovation/superstars%20the%20dynamics%20of%20firms%20sectors%20and%20cities%20leading%20the%20global%20economy/mgi\\_superstars\\_discussion%20paper\\_oct%202018-v2.ashx](https://www.mckinsey.com/~/media/mckinsey/featured%20insights/innovation/superstars%20the%20dynamics%20of%20firms%20sectors%20and%20cities%20leading%20the%20global%20economy/mgi_superstars_discussion%20paper_oct%202018-v2.ashx)
- Mckinsey (2019) Twenty-five years of digitization: ten insights into how to play it right. Briefing note prepared for the Digital Enterprise Show. Madrid, 21–23 May 2019. McKinsey Global Institute <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/twenty-five-years-of-digitization-ten-insights-into-how-to-play-it-right>
- Mercola J (2020) Google and Big Tech bough Congress. *Mercola*. 2 Jan 2020. <https://articles.mercola.com/sites/articles/archive/2020/01/07/google-big-tech-bought-congress.aspx>
- Mikijans J, Gyarmati L, Erramilli V, Laoutaris N (2012) Detecting price and search discrimination on the Internet. In: *Proceedings of the 11th ACM workshop on hot topics in networks (HotNets-XI)* October 2012. pp 79–84. <https://doi.org/10.1145/2390231.2390245>
- Möhlmann M, Zalmanson L (2017) Hands on the wheel: navigating algorithmic management and Uber drivers' autonomy. In: *Proceedings of the international conference on information systems ICIS 2017*. 10–13 December, Seoul, South Korea
- Munich D, Psacharopoulos G (2018) Education externalities: what are they and what we know. *EENEE Analytical report* No. 34. [file:///C:/Users/pedrpab/AppData/Local/Packages/Microsoft.MicrosoftEdge\\_8wekyb3d8bbwe/Temp-State/Downloads/EENEE\\_AR34%20\(1\).pdf](file:///C:/Users/pedrpab/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/Temp-State/Downloads/EENEE_AR34%20(1).pdf)
- NSF (2017) Leading cloud providers join with NSF to support data science frontiers. *National Science Foundation news release 18-009*. [https://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=244450](https://www.nsf.gov/news/news_summ.jsp?cntn_id=244450)
- Obar JA, Oeldorf-Hirsch A (2018) The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services. *Inf Commun Soc* 23:128–147. <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1486870>
- OECD (2016) The Cancun Ministerial declaration on Digital Economy: Innovation, Growth, and social prosperity. The Cancun Declaration. 23 June 2016. <https://web.archive.oecd.org/2016-06-24/407106-oecd-digital-economy-ministerial-declaration.htm>
- OECD (2019) *Digital innovation: seizing policy opportunities*. OCDE Publishing, Paris. <https://doi.org/10.1787/a298dc87-en>
- OECD (2019) Secretariat Proposal for a 'Unified Approach' under Pillar One. <https://www.oecd.org/tax/beps/public-consultation-document-secretariat-proposal-unified-approach-pillar-one.pdf>
- Pawelke A, Tatevossian AR (2013) Data philanthropy: where are we now? *United Nations Global Pulse*. 8 May 2013. <https://www.unglobalpulse.org/data-philanthropy-where-are-we-now>
- Pedraza P de, Visintin S, Tijdens K, Kismihok G (2019) Survey vs scraped data: comparing time series properties of web and survey vacancy data. *IZA J Labour Econ*, 8, 1 <https://doi.org/10.2478/izajole-2019-0004>
- Porter E (2016) With competition in tatters, the rip of inequality widens. *New York Times*. 12 July 2016. <https://www.nytimes.com/2016/07/13/business/economy/antitrust-competition-inequality.html>
- Posner E, Weyl G (2018) *Radical markets: uprooting capitalism and democracy for a just society*. Princeton University Press, Princeton, NJ
- Pratley N (2018) UK finally takes on arrogant tech giants with digital services tax. Budget levy on giants such as Facebook, Google and Amazon could go further—but it's a start. *The Guardian*. 29 October 2018 <https://www.theguardian.com/uk-news/2018/oct/29/uk-digital-services-tax-budget-facebook-google-amazon>
- Redman TC (2008) *Data driven: profiting from your most important business asset*. Harvard Business Press. ISBN 978-1-4221-6364-1
- Redman TC (2018) Data quality: treating data as an asset, a prerequisite for success in a data driven future. <https://www.dnvgl.com/feature/data-quality.html>
- Romer PM (1986) Increasing returns and long-run growth. *J Political Econ* 94(5):1002–1037
- Rowley J (2007) The wisdom hierarchy: representations of the DIKW hierarchy. *J Inf Sci* 33(2):163–180. <https://doi.org/10.1177/0165551506070706>
- Samuelson PA (1948) *Economics*. McGraw-Hill, New York
- Samuelson PA and Nordhaus WD (2010) *Economics*, 19th ed, McGraw-Hill, New York
- Sandle P (2018) Britain to target online giants with new 'Digital Services Tax'. *Reuters*, 29 October 2018. <https://uk.reuters.com/article/us-britain-budget-digital-tax/britain-to-target-online-giants-with-new-digital-services-tax-idUKKCN1N3265>
- Schroeder R (2016) Big data business models: challenges and opportunities. *Cogent Social Sciences*, 2:1. Halsall J (Reviewing Editor) <https://doi.org/10.1080/23311886.2016.1166924>
- Schroeder R, Cows J (2014) Big data, ethics, and the social implications of knowledge production <https://dataethics.github.io/proceedings/BigDataEthicsandtheSocialImplicationsofKnowledgeProduction.pdf>
- Schumpeter JA (1942) *Capitalism, socialism and democracy*. New York: George Allen & Unwin
- Scott M, Young Z (2018) France and Facebook announce partnership against online hate speech. Emmanuel Macron has teamed up with Mark Zuckerberg to review the country's regulatory response to the issue. *Politico*. 11 December 2018 <https://www.politico.eu/article/emmanuel-macron-mark-zuckerberg-paris-hate-speech-igf/>
- Scott Morton F, Bouvier P, Ezracchi A, Jullien B, Kazt R, Kimmelman G, Melamed AD, Morgenstern J (2019) Report for the study of digital platforms market structure and anti-trust subcommittee. 15 May 2019. <https://www.judiciary.senate.gov/imo/media/doc/market-structure-report%20-15-may-2019.pdf>
- Shiller BR (2014) First-degree price discrimination using big data. [https://www.brandeis.edu/economics/RePEc/brd/doc/Brandeis\\_WP58R2.pdf](https://www.brandeis.edu/economics/RePEc/brd/doc/Brandeis_WP58R2.pdf)
- Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T, Hassabis D (2017) Mastering the game of Go without human knowledge. *Nature*, Vol. 550 <https://doi.org/10.1038/nature24270>, <https://www.nature.com/articles/nature24270.pdf>
- Singer N, Isaac M (2020) Facebook to pay \$550 million to settle facial recognition suit. *The New York Times*. 29 January 2020. <https://www.nytimes.com/2020/01/29/technology/facebook-privacy-lawsuit-earnings.html>
- Statista (2015) Google's annualized advertising ARPU from the 1st quarter of 2012 to the 1st quarter of 2014 (in US dollars) <http://www.statista.com/statistics/306570/google-annualizedadvertising-arpu/>. Accessed 7 Dec 2015
- Steel E (2013) Companies scramble for consumer data. *Financial Times*. 12 June 2013. <https://www.ft.com/content/f0b6edc0-d342-11e2-b3ff-00144feab7de>
- Steel E, Locke C, Cadman E, Freese B (2013) How much is your personal data worth? Use our calculator to check how much multibillion-dollar data broker industry might pay for your personal data. *Financial Times*, 12 June 2013. <https://ig.ft.com/how-much-is-your-personal-data-worth/>. Accessed 28 November 2018
- Stempel J (2020) Google faces \$5 billion lawsuit in U.S. for tracking 'private' internet use. *Reuters*. 2 June 2020. <https://www.reuters.com/article/us-alphabet-google-privacy-lawsuit/google-faces-5-billion-lawsuit-in-u-s-for-tracking-private-internet-use-idUSKBN23933H>
- Stiglitz JP (2001) Information and change in the paradigm in economics. *Nobel Prize lecture*, 8 December <https://www.nobelprize.org/prizes/economic-sciences/2001/stiglitz/lecture/>
- Taylor L, Schroeder R, Meyer E (2014) Emerging practices and perspectives on big data analysis in economics: bigger and better or more of the same? *Big Data Soc*, 1–10
- Tett G (2018) Recalculating GDP for the Facebook age. The true impact of social media? Economists are approaching the question from a different angle. *Financial Times*. <https://www.ft.com/content/93ffec82-ed2a-11e8-8180-9cf212677a57>
- The Economist (2017) Fuel of the future: data is giving rise to a new economy. *The Economist*. 6 May 2017. <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy>
- The Economist (2018) American tech giants are making life tough for startups. 2 June 2018. <https://www.economist.com/business/2018/06/02/american-tech-giants-are-making-life-tough-for-startups>

- Toscano J (2019) China will outpace AI capabilities but will it win the race? Not if we care about freedom. *Forbes*. 3 December 2019. <https://www.forbes.com/sites/joetoscano/2019/12/03/china-will-outpace-us-artificial-intelligence-capabilities-but-will-it-win-the-race-not-if-we-care-about-freedom/#32e29da66ed3>
- Uber (2018) Fraudulent trips: how to recognise fraud. <https://www.uber.com/en-ZA/drive/resources/recognising-fraud/> Accessed January 2019
- Ulloa J (2019) 'Newsom wants companies collecting personal data to share the wealth with Californians', *Los Angeles Times*, 5 May <https://www.latimes.com/politics/la-pol-ca-gavin-newsomcalifornia-data-dividend-20190505-story.html>
- UN (2013) A new global partnership: eradicate poverty and transform economies through sustainable development, United Nations Publications. <https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=893&menu=1561>
- UN (2014) A world that counts: mobilising the data revolution for sustainable development. Independent Experts Advisory Group on Data Revolution for Sustainable Development, November 2014, United Nations. <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>
- Unctad (2017) World investment report 2017, investment and the digital economy, UN Publications, Geneva
- Ursu MR (2018) The power of rankings: quantifying the effect of rankings on online consumer search and purchase decisions. *Mark Sci* 37:530–552. <https://doi.org/10.1287/mksc.2017.1072>
- Varian HR (2013) Big data: new tricks for econometrics. *J Econ Perspect* 28(2):3–28
- Varian H (2018) Artificial Intelligence, Economics, and Industrial Organization, NBER Working Paper 24839. <https://www.nber.org/papers/w24839>
- Vaughan R, Hawksworth J (2014) The sharing economy: how will it disrupt your business? Megatrends: the collisions. PriceWaterhouseCoopers, London, [https://pwc.blogs.com/files/sharing-economy-final\\_0814.pdf](https://pwc.blogs.com/files/sharing-economy-final_0814.pdf)
- Vigo R (2013) Complexity over uncertainty in generalized representational information theory (GRIT): a structure-sensitive general theory of information. *Information* 4:1–30. [http://cogprints.org/8784/1/Vigo%20\(2013\).pdf](http://cogprints.org/8784/1/Vigo%20(2013).pdf)
- WEF (2019) Global Technology Governance A Multistakeholder Approach. World Economic Forum White Paper. 1 November 2019. <https://www.weforum.org/whitepapers/global-technology-governance-a-multistakeholder-approach>
- Wheeler T (2021) Time for U.S.-EU digital alliance, *Brookings* Thursday 21 2021 <https://www.brookings.edu/articles/time-for-a-us-eu-digital-alliance/>
- White House (2015) Big data and differential pricing. White House Council of Economic Advisers [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/docs/Big\\_Data\\_Report\\_Nonembargo\\_v2.pdf?utm\\_source=Bruegel+Updates&utm\\_campaign=656e7da39b-Blogs+review+11%2F02%2F2017&utm\\_medium=email&utm\\_term=0\\_eb026b984a-656e7da39b-278510293](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf?utm_source=Bruegel+Updates&utm_campaign=656e7da39b-Blogs+review+11%2F02%2F2017&utm_medium=email&utm_term=0_eb026b984a-656e7da39b-278510293)
- Wiener M, Saunders C, Marabelli M (2020) Big-data business models: a critical literature review and multiperspective research framework. *J Inf Technol* 35(1):66–91. <https://doi.org/10.1177/0268396219896811>
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3:160018. <https://doi.org/10.1038/sdata.2016.18>
- Wong JC (2017) 'Greyball: how Uber used secret software to dodge the law', *The Guardian*, 4 March <https://www.theguardian.com/technology/2017/mar/03/uber-secret-program-greyballresignation-ed-baker>
- Worstell T (2017) Understanding the economic value of your personal data. *www.computerweekly.com* 26 May 2016 <https://www.computerweekly.com/opinion/Understanding-the-economic-value-of-your-personal-data>
- WTO (2018) World Trade Report 2018: The future of world trade. How digital technologies are transforming global commerce. [https://www.wto.org/english/res\\_e/publications\\_e/wtr18\\_e.htm](https://www.wto.org/english/res_e/publications_e/wtr18_e.htm)
- Zhang B, Kreps S, McMurry N, McCain RM (2020) Americans' perceptions of privacy and surveillance in the COVID-19 pandemic. *Plos One*. 15(12):e0242652
- Zuboff S (2019) The age of surveillance capitalism: The fight for the future at the new frontier of power. Profile Books Ltd. 9781781256855

## Acknowledgements

The authors would like to acknowledge comments and suggestions from Bertin Martens, Daniel Vertesy, Michaela Saisana, Sven Langedijk, and Ignacio Sanchez and previous work and discussions conducted at Webdatanet Task Force 25 during the cost action webdatanet (IS-1004, <https://webdatanet.usal.es/>), at a series of seminars held at the European Commission Joint Research Centre in Ispra, Italy, and at the 2020 CARMA Conference on Advanced Research Methods and Analytics <http://archive.carmaconf.org/carma2020/index.html>. The scientific output expressed does not imply a European Commission policy position. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

## Author contributions

Both authors have contributed to the conception of the work and revising it critically. Both authors have approved the final version to be published. They agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Competing interests

The authors declare no competing interests.

## Informed consent

Informed consent was not necessary because this article does not report any studies with human participants performed by any of the authors.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-023-02096-w>.

**Correspondence** and requests for materials should be addressed to Pablo de Pedraza.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023